

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>H04M 3/22</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/53589</b> <b>(43) International Publication Date:</b> 26 November 1998 (26.11.98)
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%; vertical-align: top;"> <p><b>(21) International Application Number:</b> PCT/GB98/01305</p> <p><b>(22) International Filing Date:</b> 6 May 1998 (06.05.98)</p> <p><b>(30) Priority Data:</b>            97303375.6 16 May 1997 (16.05.97) EP  <i>(34) Countries for which the regional or international application was filed:</i> GB et al.            9726125.9 10 December 1997 (10.12.97) GB</p> <p><b>(71) Applicant (for all designated States except US):</b> BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).</p> <p><b>(72) Inventors; and</b>  <b>(75) Inventors/Applicants (for US only):</b> HOLLIER, Michael, Peter [GB/GB]; Grange Farm, 15 Herbert Road, Kesgrave, Ipswich, Suffolk IP5 2XX (GB). REYNOLDS, Richard, John [GB/GB]; 103 Oaksread Close, Ipswich, Suffolk IP4 4HW (GB). RIX, Antony, William [GB/GB]; Bottom Flat, 164 Huntingdon Road, Cambridge CB3 0LB (GB).</p> <p><b>(74) Agent:</b> LIDBETTER, Timothy, Guy, Edwin; BT Group Legal Services, Intellectual Property Dept., Holborn Centre, 8th floor, 120 Holborn, London EC1N 2TE (GB).</p> </div> <div style="width: 48%; vertical-align: top;"> <p><b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b>  <i>With international search report.</i></p> </div> </div>		
<b>(54) Title:</b> TESTING TELECOMMUNICATIONS EQUIPMENT		
<pre> graph TD     20((Telephone network)) &lt;--&gt; 21[QA Node]     20 &lt;--&gt; 22[QA Node]     21 --&gt; 23[Network management system]     22 --&gt; 23       </pre>		
<b>(57) Abstract</b> <p>A communications system is tested to simulate a realistic conversation. A first measurement device (21) makes a call, through the system (20) it is testing, to a second measurement device (22), and the devices (21, 22) converse using predetermined speech generated dynamically by each device in response to the signals received from the other, to simulate aspects of conversation over non-perfect communications system link. The progress of the conversation will therefore take different courses according to the quality of the link (20). For example, if an expected signal is detected by one device (22), it transmits an appropriate predetermined response to device (21). However, if the expected signal is corrupted, or not received at all, different predetermined responses (or no response at all) are transmitted to the device (21). This in turn may cause the device (21) to repeat the original signal. Because all the responses, including those to poor quality signals, are predetermined, each device (21, 22) can readily determine the quality of the signals received, because each signal received by one device from the other must be one of a limited number of possible responses to the last transmission made by the one device to the other. The devices (21, 22) perform measurements on the sounds they receive from each other, and monitor the progress of the conversation to compute parameters describing aspects of the received signal quality.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**TESTING TELECOMMUNICATIONS EQUIPMENT**

This invention relates to the quality assessment of communications systems.

5 In the increasingly liberalised telecommunications market, differentiation by quality is an important factor. The widespread installation of coding equipment in recent years has created the need for a new generation of assessment techniques. Conventional analysis assumes that the system is linear and time-invariant, and characterises it on the basis of delay, frequency response, noise  
10 level and noise spectrum. Modern networks exhibit far more complex effects and contain elements such as speech switches and compressive codecs that are highly non-linear and time varying.

Two main techniques are used in the industry to characterise the subjective quality of networks and their components. In listening tests a panel of  
15 subjects hear a series of sound clips that have been passed through simulated network conditions. Conversational tests require pairs of subjects to communicate through several simulated telephone links. In both cases the subjects are commonly asked to vote on a five point scale, from 'excellent' to 'bad'. Averaging across all subjects produces a mean opinion score (MOS) for each condition that  
20 mirrors the subjective quality of the network while reducing the random errors that appear in subjective voting.

The need to use a number of subjects for this kind of assessment makes the techniques expensive and hinders or prevents their application in the 'live' telephone network. It is therefore desirable to produce a tester that will  
25 automatically measure quality on the same scale, and that produces the same scores, within experimental error and subjective expectation, as the mean of several conventional subjective tests. As will be discussed later, assessment devices have been developed that can predict the listening quality of speech passing one way through the telephone network, and have been extended by using  
30 a variable speech level to estimate the conversational quality of a two way link.

The current state of research in this area is to perform a number of measurements on the communications system under test, such as echo, delay, or degradation of speech, and predict the system's quality from these measurements.

However, these tests neglect the reactions of users to the system's behaviour, which in turn can influence the way the system performs.

Telecommunications companies have considerable experience in using human subjects to assess the subjective quality of a communications network.

5 This knowledge has been built up through international groups, such as ETSI and the ITU: see, for example, *Methods for subjective determination of transmission quality*: ITU-T Recommendation P.800. Subjective assessment uses a panel of subjects, who vote on a number of candidate network conditions. Their votes are averaged and examined, and give information about the subjective quality of the  
10 networks. This method is valued because the scores are directly related to peoples' opinion of quality. In contrast, conventional engineering metrics such as signal-to-noise ratio do not in general correlate well with speech quality. The subjective tests relevant to the present invention may be classified as:

listening, where the subjects hear sections of speech that have been  
15 passed through test networks and vote on what they hear, and

conversational, where two subjects talk to each other over candidate network connections and vote on the speech quality of each conversation.

Different questions prompt the subjects to vote on different aspects of their perception of the network, such as effort or quality.

20 Techniques have been developed (*Models for predicting transmission quality from objective measurements*: ITU-T P series recommendations: Supplement 3) to estimate the conversational quality of a conventional, linear, network using classical signal processing metrics such as echo delay and level. These measures rely on analysis of the network using artificial signals — usually  
25 sine waves or noise bursts — which may not be passed by a communications network designed for speech.

Models of human perception allow a more sophisticated analysis of the quality of speech transmitted by the network. Key features of the human senses, such as masking and threshold of hearing, are used to deduce whether errors in  
30 transmission are audible, and estimate their subjectivity. Higher level 'perceptual' processing takes further account of the amount and distribution of errors, and their coincidence with certain parts of speech. This computation allows the subjective mean opinion score that would be given by a series of subjective tests to be

predicted. Such models are described in: ITU-T Recommendation P.861: *"Objective quality measurement of telephone-band (300-3400 Hz) speech codecs"*, and International Patent Applications WO94/00922 and WO95/15035.

The methods of conversational assessment described above may be used  
5 to weight perceptually-motivated quality measures to derive a perceptually-based estimate of conversational score. A system is described in the present Applicant's co-pending International Patent application (having the same filing date and claiming priority from the same two applications as the present application), which varies vocal level until an equilibrium is achieved, estimating listening effort, and  
10 then uses a conversational weighting based on echo and delay to estimate the quality of the connection for conversational speech.

There are good reasons for using artificially-generated speech-like signals, rather than recordings of human talkers, for testing a network designed to carry speech. In particular, artificial speech can be constrained to contain precisely  
15 defined phonemes, and are more easily reproducible to allow comparison of results. Artificial speech suitable for conversational testing is described in International Patent applications WO94/00922 and WO95/01011.

A dynamic conversational tester could take account of the change in certain properties of speech in hostile conditions. A description of these properties  
20 is given in International Patent application WO96/06495.

At the start of a call during which network quality is to be assessed, knowledge of certain network characteristics, such as delay and echo, is unavailable. However, these characteristics affect the conversational pattern. Furthermore, a fixed conversational pattern does not take into account the human  
25 response to losses in the line, such as noise, delay or freeze-out (the loss of a signal, for example due to capacity problems in the network, or delays in response to the start of a signal), which may be random in nature.

According to the invention there is provided a method of testing a communications system in which a first measurement device makes a call, through  
30 the system it is testing, to at least one second measurement device, and the devices converse using predetermined speech signals, and at least one of the devices performs measurements on the sounds it receives from the other device or devices, to compute one or more parameters describing aspects of the received

signal quality, wherein the predetermined speech signals generated by each device are generated dynamically in response to the signals received by said devices, to reproduce aspects of conversation over a non-perfect communications system link.

This allows a dynamic response to events such as speech loss, mimicking  
5 the behaviour of humans communicating in the same conditions. Behaviour such as stand-off and recovery after loss, which will be described later, can also be incorporated. This allows the communications link to be evaluated in a more representative way than before. The signals may be assessed using a number of techniques to obtain measures that more accurately reflect the conversational  
10 quality of the system. In particular the quality of the communications system link can be determined in accordance with the information content and/or quality of the response signals received, and the actual progress of the simulated conversation.

In a preferred arrangement a device which does not clearly receive an expected signal from another device responds by transmitting a predetermined  
15 reply signal which, if received uncorrupted by the other device, will cause the other device to repeat the expected signal. (Of course, this reply may itself be corrupted in transmission, causing the other device in turn to respond in some other way than this intention).

Some of the signals to be generated for transmission in response to the  
20 input signals may be selected according to the quality of the received signals, and some may be signals requesting repetition of a received signal, selectable in response to poor quality of the received signal.

Properties such as the volume or speed of the response signals to be generated may be selected according to the quality of the received signals.

25 The individual devices may transmit and recognise device-specific identification signals forming part of the speech generated, such as spoken words, thus allowing the individual path under test to be identified.

The method may incorporate a system delay calculation process in which a first signal is transmitted from a first device to a second device, the second  
30 device receives the first signal and, after a predetermined delay, transmits a response signal to the first device, and the system delay is determined from the time of transmission of the first signal, the time of receipt of the second signal, and the predetermined delay. This process measures round trip delay. It is also

possible to measure one-way delays between two of the devices having synchronised clocks, wherein one of the devices transmits to another device a signal indicative of the current time, and the delay in the system under test is determined by comparing the received time-indicative signal with the time of receipt.

The signals used may be pre-recorded samples of human speech or artificially-generated speech samples, representative of a plurality of different talkers.

In a preferred arrangement, one of the devices may transmit a signal simultaneously with another signal being received by the same device, such that corruption of one or other of the signals at its respective point of receipt can be detected.

The invention may be used for assessing the conversational quality of telephone and teleconference networks. It may be used on laboratory models of the networks, for example in the development of new equipment or in the planning of new networks, or for the assessment of a 'live' network, either installed at the customer premises, the local exchanges, or, via digital communications links, at central objective measurement site. Finally, the inclusion of a model of conversational state in the transmission system of the network would allow the network to deliver conversationally optimised speech under constraints such as limited bit rate.

In the case of a laboratory or planning assessment, the quality predictions would be used to support design decisions, probably in conjunction with a limited number of subjective tests. Assessment equipment installed in the network could be used to make test calls to monitor the quality of the core network, to assess the quality of a competitor's or any other interconnecting network, or to guarantee the quality of calls provided to a customer. Such equipment would pass results to a central monitoring location; in the case of malfunction it could raise alarms and provide diagnostic information to assist the rectification of faults.

Applications of conversational assessment are not limited to conventional telephony. The same techniques could be applied to other voice transmission systems, such as Internet phones or audio conference systems. Furthermore, the techniques could be employed to operate on the audio component of videophone,

audio/visual teleconference systems, or shared communications environments such as virtual worlds.

It is envisaged that a model of the conversational state of two human talkers could be used in the network delivery infrastructure to ensure that, under  
5 constraints such as limited data rate, conversational quality could be optimised for a number of customers.

According to another aspect, there is provided a device for simulating a conversation over a communications system, comprising:

- a store for storing a plurality of signals forming a conversation
- 10 a receiver for receiving signals from one or more complementary devices,
- a comparison device for comparing signals received by the receiver with signals stored in the store;
- selection means responsive to the comparison device for selecting from the store a signal for transmission to the complementary device or devices;
- 15 transmission means for transmitting a signal selected by the selection means to the complementary device or devices.

A measurement system can be made up of any number of such devices. Two devices can be used to generate a simulated two-party conversation over a network. Three or more can be used to simulate a conference call. In laboratory  
20 conditions, it is possible for a single such device to act as both/all ends of the conversation, by connecting it to both/all ends of the equipment under test. In these circumstances, the device is acting as its own "complementary device", in the sense used above.

One of the devices has conversation initiating means for generating an  
25 initial signal for transmission by the transmission means. One or more of the devices have processing means for measuring the conversational quality of the communications system. The processing means may include means for monitoring the operation of the selection means to assess the progress of a conversation, and/or means for monitoring the quality of signals received from the  
30 complementary devices by the receiving means.

The selection means preferably has means to select different signals for transmission according to which of the predetermined signals is received, and according to the quality of the received signals. This means may be arranged to



select a signal requesting repetition of a received signal in response to poor quality of the received signal

The selection means may also have means to select properties of the signal to be transmitted according to the quality of the received signal.

- 5       The processing means may include means for monitoring the progress of the simulated conversation, and determining the quality of the communications system link thereby.

One of the devices may include delay calculation means comprising means for determining the time at which a first signal is transmitted to a complementary  
10 device, means for determining the time at which a predetermined response signal is received from the complementary device, means for retrieving from the store the value of a predetermined delay at the complementary device between receipt of the first signal and transmission of the response signal, and means for determining therefrom the additional delay imposed by the system under test.

- 15       Two or more of the devices may have synchronised clocks, wherein at least one of the devices has means for transmitting a signal indicative of the current time, and at least one other of the devices has means for determining the delay in the system under test by comparing the received time-indicative signal with the time of receipt.

- 20       The devices may have means for simultaneously transmitting and receiving predetermined signals, and means for detecting whether the predetermined received signal is corrupted by the presence of the transmitted signal.

- The key measurement to be made on the network under test is the quality of conversational speech. This measurement depends on a number of different  
25 factors, such as:

- subjective quality of speech;
- side tone (the deliberately-imposed electrical coupling normally provided between the mouthpiece and earpiece of a telephone)
- degradation of speech due to filtering and coding in each direction;
- 30 difficulty in communication caused by network characteristics such as:
  - talker echo (the unwanted return of a signal to its point of origin, due to electrical mismatches in the system, or acoustic feedback at a remote handset),

listener echo (multiple, temporally spaced, arrivals of the same signal),  
and

freeze-out (failure to transmit part of a signal due to non-availability of  
the channel, delayed action of voice activity detectors, etc);

5       The assessment of these network characteristics requires two steps;  
firstly sending speech over the network, and secondly, receiving the speech signal  
at each termination and determining the subjective impact of the observed  
degradations in the received signal.

10       In this invention, the speech sent across the network is varied dynamically  
in response to the received signals using a conversational model.

      However, because both devices have a predetermined set of responses to  
any signal (the response to be selected depending on what signal is received and  
the quality of that signal) the responses received by one device are indicative of  
the signal quality received by the other and the conversation is constrained, such  
15   that comparative measures are possible.

      More specialised measurements may be performed at the same time to  
allow different degradations to be characterised for network diagnostics. The  
recordings of received speech, and knowledge of the speech that was sent (which  
is also available from the conversational model) are sufficient for these  
20   measurements, which include:

      echoes from the local and the far end of the connection, or any  
intermediate points;

      degradation due to the (time-varying) operation of echo cancellers;

25   loss or gain in speech loudness, and the variation of this loss with time  
and speech level;

      clipping due to the imperfect operation of speech detectors, and the  
variation of clipping with speech level;

      freeze-out (also known as dropout) due to instantaneous overloading in  
certain channels - a particular example is observed when intervals between speech  
30   are used to carry data packets: the beginning of an utterance may be lost as the  
system completes transmission of a packet;

      noise, which may be due to electrical sources, or crosstalk with data or  
speech signals;

inability of the system to handle "double-talk" (the transmission of speech in both directions simultaneously).

The network under test is likely to contain elements specifically designed to operate with speech. For example, speech codecs are trained to code the  
5 different parts of speech well, but may behave erratically when passing non-speech signals such as tones or white noise. A further point of note is that the presence of non-speech signals such as tones may alter the properties of the system. In extreme cases certain tones may switch off echo cancellers or cause the system to clear the channel through for data traffic.

10 It is therefore advantageous that a system for assessing the quality of such a network for passing speech should use only speech or speech-like signals in its tests. This is not to say that the network should not be assessed for its ability to transmit tones (such as DTMF digits) and data, merely that this is a separate problem.

15 Certain properties of conversational speech depend on the quality of the transmission network. Vocal level and vocal tone change as a response to level or degradation of received speech. The conversational pattern is affected by characteristics such as delay, resulting in behaviours such as increased likelihood of double-talk and "false-starts", which occur when both parties start talking at  
20 about the same time, and one or both of them stop on realising that the other is also talking. Human talkers also go through processes of resynchronisation after speech has been lost or a false-start made. These properties may be included in the conversational model to ensure that artificial testing of the network is performed in a representative way.

25 Other aspects of variability in speech are significant in that they may induce different behaviour in the network and hence be degraded differently in transmission. Aspects of culture and talker characteristics may also cause changes in system performance. Although these effects are more complex than those listed above, their impact on subjective quality makes them worth considering. Example  
30 effects include: talker gender; language, grammar and vocabulary; formal or informal speech; information content — whether it is important that the speech is understood correctly; and talker assertiveness — determining the nature of the variation of vocal level with received level.

Account may be taken of these classes of variability by using several different conversational models to simulate different talkers and languages, to gain an overall quality assessment.

A number of the network characteristics described above may vary with time. For example, echo cancellers take a certain period of time to train; during this  
5 training period unexpected and confusing sounds may be heard, and account must be taken of this temporal variability and the higher-level perceptual implications of where in the conversational pattern the degradations occur.

An embodiment of the invention will now be described, with reference to  
10 the Figures, in which:

Figure 1 is a schematic representation of a quality assessment system,

Figure 2 is a schematic representation of the general arrangement of a conversational assessment device according to the invention,

Figures 3 and 4 are flow diagrams illustrating the processes respectively  
15 performed by two such devices simulating a conversation,

Figure 5 illustrates the progress of such a conversation in ideal conditions,

Figure 6 illustrates the progress of such a conversation in the presence of various line degradation conditions.

An outline structure for a quality assessment (QA) system with two  
20 devices is shown in Figure 1. As shown in Figure 1, at each termination 21, 22 of the network 20 under test there is located a measurement device or tester. One device (the 'originator') 21 sets up calls as appropriate and initiates a conversation with one or more receiving devices ('destinations') 22 (only one shown); assessment of an audioconference link would require one such device for each  
25 party to the call. Once conversation is initiated all devices 21, 22 etc behave in a similar way.

During the call the devices 21, 22 communicate through the network 20 using speech-like signals. The signals sent by each device are governed by a model of the conversation, which determines what is 'spoken' based on what has been  
30 received and an underlying state encoding the desired content ('conversational intent') of the call. Quality assessment is performed by at least one device, and preferably by all of them, based on the incoming speech and on knowledge of the signals that would have been sent by the other device or devices. It is possible but

not necessary that this assessment be performed in real time, as the speech is received, but the received speech may be recorded and assessed subsequently. In order to obtain a comprehensive assessment of the quality of the link as a whole, the results of assessment carried out by each device can be transmitted to each other or to a network management centre 23, in order to combine the various measurements. The network management centre may display the quality measures to an operator, or may be arranged to act on the measures automatically, for example to re-route a call on which low quality is identified, or to alert maintenance staff.

- 10        Each device 21, 22 in a conversation has its own state: that is, a predetermined set of responses to given stimuli, which will vary according to which stimuli are received and the perceived quality of these stimuli. As shown in Figure 2, each device 21, 22 has knowledge of the state (51, 61) of itself and the other device or devices. During a perfect conversation (with no loss or delay) each
- 15        device 21, 22 would speak a pre-determined sequence of phrases or words, in response to phrases received from the other device. The state is defined more precisely to represent whether the device is listening, and the speech it will generate in different circumstances. After losses on the line the device will take steps to recover by performing a spoken re-synchronisation with the other devices.
- 20        The state's variation ('state trajectory') is determined by a comparison between what is received and the expected signal, and by a pre-determined desired sequence of speech to pass between the devices ('conversational intent'). The model is therefore abstracted into two parts, a state (the current action) and a processor, which determines transitions in the state.
- 25        The comparison between received and expected speech is performed using an analysis of similarity such as the perceptual models previously described, taking into account factors such as echo, noise, and filtering of the signal. If the speech contains the expected sequence of sounds, within a defined amount of loss, the device either begins talking or continues to listen, dependent on the pre-determined
- 30        conversational intent.

      If something unexpected is heard (such as distortion, speech loss, or a predetermined recovery signal from another device or devices), recovery behaviour is initiated, again using a pre-determined sequence of sounds. The recovery

behaviour may be recursive, continuing throughout a sustained loss in channel quality. The fact that recovery has been carried out may change certain features of the future conversational content of the call. The characteristics of received speech will also determine certain parameters of generated speech (such as vocal  
5 level and tone).

To assist in dealing in processing incoming signals, the speech communicated by each device may be labelled with a specific identity, just as human talkers can be identified from their speech. Devices may also identify themselves more clearly by name or a recognised identity sequence.

10 A model of conversational state as has just been described can be applied to the assessment of the conversational quality of a network as follows. One or more devices set up a call in which they may communicate with each other. Each has a different sequence of information to send to the other devices, and has prior knowledge of what it should hear in reply. The devices converse using a model of  
15 conversational state that allows them to communicate this information, and react to noise, echoes or loss on the line, in a predetermined way that is representative of human talkers. The key invention in this technique is this dynamic variation of speech, allowing the received signal to change what is to be generated in response. This allows a network connection to be realistically exercised to identify  
20 whether it would impair conversation, and rated accordingly.

The information (or desired speech sequence) to be communicated forms an idealised conversational pattern for each talker. This should contain a representative sample of different parts of speech, and reproduce the temporal structure and content of a typical conversation. The sequence should exercise  
25 conversational features such as interruption, double-talk and hand-over (the correct exchange of listening and talking functions between the two users). Knowledge of the quality of the line so far may be included in the information to be communicated, as will be illustrated in the worked example. Finally, the amount of spoken material to be used should also be representative.

30 Each device uses its own conversational state, and knowledge of the conversational state of the other device as acquired during the conversation, to determine what signals it is to send. The conversation will therefore exercise the system under test in a way that is representative for the system under test. The

measurements described previously are performed, computing objective predictions of subjective parameters such as quality, as well as a number of other objective quantities such as echo characterisation.

A functional diagram of such a device is given in Figure 2, using a layered  
5 form that separates the algorithmic model 30 from the network interface functions 40 for connection to a network 49, which may be 2- or 4-wire telephony, digital transmission systems such as ISDN or T1 or any other suitable system. The key elements of the device are:

the models of the talker 50 and other talkers 60, incorporating  
10 conversational state models 51 and 61, controlling speech generation processes 52 and 62;

perceptual analysis of the received speech 70, measured against sent and expected received speech;

a conversational processor 80 which controls the conversation states 51,  
15 61 to determine the state trajectory and applies both the conversational intent (speech to be communicated in normal conversation) and recovery behaviour after losses, double-talk or other unexpected events on the line.

To discuss Figure 2 in more detail, the algorithmic model 30 comprises the talker models 50, 60 which model the conversation state of each talker in the  
20 system (conversation state units 51, 61) in order to generate the speech for transmission (speech generation unit 52), and the speech which is expected to be received from the remote user (speech generation unit 62). The speech generated by the speech generation unit 52 is transmitted through a voice channel unit 31 to the network interface 40, through a corresponding voice channel unit 41, for  
25 transmission to the remote user. Signals received from the remote user are received from the network interface 40 and received through the voice channel 31 as received speech, and is processed by the perceptual analysis unit 70, as well as by a unit 32 which handles conventional channel measurements such as echo detection. In the perceptual analysis unit 70 the received speech is compared with  
30 the expected speech generated by the unit 52 in order to analyse the quality of the received speech, the results of which are processed by the conversational processor 80, which in turn updates the conversation states 51 and 61 in talker models 50 and 60. The conversational processor 80, and other channel

measurements 32, control a call management reporting unit 33, which handles other functions such as fax or modem test calling (module 34) and the initial voice call dialling and set up functions 35, which interface with the network through an application processor interface 36.

5       The network interface 40 has a direct voice channel interface 41 for receiving signals from the voice channel 31 in the processor 30, and a call signalling and data protocol data handler 46 for interfacing with the application processor interface 36. These interfaces 41, 46 co-operate with a low level software driver 47 and thus with the physical connection 48 with the network  
10 interface 49.

A simple example of the conversational assessment of a two-device link will now be described, with reference to Figures 3, 4, 5 and 6. Figures 3 and 4 are simplified flow charts showing how the conversation would progress in various circumstances as seen by the originator unit 21 and receiving unit 22 respectively.  
15 These flow diagrams merely illustrate the principles employed, and have been very much simplified from what would actually be required to emulate all possible fault conditions realistically. In particular, each step shows only two possible outcomes, but in practice there may be several.

Figure 3 shows the processes carried out by the device 21, and Figure 4  
20 the processes carried out by the device 22, when interacting with each other. As can be seen, the conversation will take different courses according to the quality of the signals received by each device from the other. Certain steps, shown in bold outline, are indicative of points at which quality measures may be made. It should be noted however that all received messages may also be monitored generally for  
25 signal quality.

As shown in Figure 3, the device 21 (which has the spoken identification signal "Smith") initiates a call (step 100) to the device 22 (which has the spoken identification signal "Jones"). The device 21 then waits to receive an expected message 1 from the device 22. In the next step, step 101, if the expected  
30 message is not received correctly, a message 12 is transmitted (step 102) in order to prompt the remote device 22 to resend the message. If message 1 is received correctly the device 21 transmits a second message 2 (step 103) and awaits a response from the remote device 22. The device 21 expects one of two responses



(messages 3/13) from the remote device 22. The two messages are indicative of the signal quality received at the remote device 22. When this message is received (step 104) the signal quality of that message is assessed (step 105). Note that this step measures the quality of the signal received by device 32, independently of its  
5 content. The content of the signal (i.e. whether it is signal 3 or signal 13) is indicative of the quality received by the remote device 22. This may also be monitored at the device 21 by identifying which of the two signals 3/13 is actually received.

In response to the measured signal quality, a message 4 or 14 is  
10 transmitted (steps 106, 107) according to whether the signal quality perceived in step 105 is good or bad. This is arranged to interrupt message 3. This allows the system's response to "double-talk" - both parties speaking at once - to be monitored. A further message is then received from the remote unit 22 (message 5, step 108) in response to which the unit 21 transmits a further message 6 (step  
15 109). In the next stage, a further message 7/15 is received (step 110). Its quality is then measured (step 111) in a similar manner to the step 105 already described and a message 16 is transmitted (step 112) if the quality of message 7/15 received in step 110 is poor. Following the transmission of message 16, the device returns to step 108 to await a repeat of message 5. However, if the quality of the  
20 signal received in step 110 (as assessed in step 111) is good, then a further step (113) is made to assess which of messages 7 or 15 was received in step 110. If the message received was message 15, the process returns to step 109, the transmission of message 6. However, if the message received in step 113 is message 7, then the system pauses (step 114) before transmitting a further  
25 message 8 (step 115), and receiving a complementary message 8 from the other system 22 (step 116). If no such message is received, the system awaits the sound of the other system hanging up (step 117) before itself hanging up (step 118). However, if message 8 is received, at step 116, the hang up process (118) happens straight away.

30 Figure 4 shows the complementary processes happening in the device 22. The device 22 initially receives the call initiation from the device 21 (step 200) in response to which it transmits message 1 (step 201) and receives in response either message 2 or message 12 from the device 21 (step 202). The content of

the signal (2/12) is indicative of the quality received by the remote device 21. This may be monitored at the device 22 by identifying which of the two messages 2 or 12 is actually received.

If message 12 is received, the device 22 re-transmits message 1 (step 201). However, if message 2 is received, the signal quality is assessed (step 203) and, according to the quality received, either a message 3 (step 204) or message 13 (step 205) is transmitted. The device 22 then receives, from the device 21, a message 4 or 14 generated by steps 106 or 107 above (step 206). The system then pauses (step 207) before transmitting a further message 5 (step 208) and the device 22 then receives from the device 21 one of two messages 6 or 16 (step 209). If the received message is message 16 then the device returns to step 208 to re-transmit message 5. However, if the message received is message 6, a quality assessment (step 210) is made. If the quality is assessed as bad, the device 22 transmits a message 15 (step 211) before returning to step 209 to await message 6 or 16 as appropriate from the device 21. However, if the quality measured in step 210 is assessed as good, the device 22 transmits a message 7 to the device 21 (step 212) and then pauses (step 214) before carrying out a process analysis as described above for concluding the conversation (steps 215, 216, 217, 218, corresponding to steps 115, 116, 117 and 118).

In the following discussion of the process, the originator 21, named Smith, is denoted by bold text, and the destination device 22, Jones, by italic text. The information to be communicated — the expected conversation for a perfect link — is shown in Figure 5 and summarised as follows: these steps are shown by double-lined arrows in Figures 3 and 4.

25

Smith makes call (step 100).

Jones answers. *Hello, Jones here* (message 1).

30

**Hello Jones, Smith speaking. How good is this line's clarity?** (message 2)

*Excellent for me* (message 3).

(Smith interrupts at start of 'me') and very good for me (message 4).

(Pause.) (step 207)

*What time is it?* (message 5)

5

**Fourteen hours fifteen minutes and ten point three seconds** (message 6).

*Thanks, that's it* (message 7).

10

(Pause.) (step 114/214)

(Simultaneously) *goodbye Smith. goodbye Jones* (message 8).

(Smith and Jones both hang up.) (step 118/218).

15

For a clear, instantaneous connection, the conversation would be as described above. In the case of a poorer quality line, with delay and heavy loss, the conversational model could give a conversation similar to the following example, illustrated in Figure 6: speech lost by the channel (sent but not received)

20 is marked below by overstriking and by 'x' in Figure 6:

Smith makes call. (Step 100)

Jones answers.

25

*Hello, ~~Jones~~ here.* (message 1)

(Smith responds to loss) **sorry, I didn't catch your name.** (message 12)

30

*Jones here.* (message 1 repeated)

**Hello Jones, Smith speaking.** How good is this line's clarity? (message 2)

*Bad for me.* (message 13)

(Smith's interruption at start of 'me' causes its loss) **and bad for me**  
(message 14).

5

(Pause.) (step 207)

*What time is it?* (message 5)

10 

**Fourteen hours fifteen...** (message 6)

(Jones interrupts)

15 

*how many hours?* (message 15: double-talk causes loss of start of  
message)

**Sorry?** (message 16)

20 

(Jones repeats question.) *What time is it?* (message 5)

**Fourteen hours fifteen minutes and ten point three seconds.** (message 6)

*Thanks, that's it.* (message 7)

25 

(Pause.) (steps 114/214)

*goodbye Smith.* (message 8) ~~**goodbye Jones**~~

30 

(Delay between devices, and network double-talk behaviour, causes loss).

Smith hangs up first (step 118);

Jones hears this (step 217) instead of the expected message 8 ("goodbye Jones"), then also hangs up (step 218)

This example illustrates the principal aspects of dynamic conversational testing. Loss of key information initiates a recovery. The first time this happens (message 1: 'Jones' lost by the channel) the recovery (step 102) gets the expected information across. The second time key information is lost ('Fourteen') (step 209), the initial recovery phrase 15 is itself garbled and the conversational state is set back (step 111), leading to a repeat of the original question (message 5: step 208). Several non-essential parts of speech are also lost; these do not initiate recovery but are taken into account in the overall assessment of signal quality.

The device 21 commences the transmission of message 4 (step 106/7) before message 3/13 (step 104) is complete. This allows the device to detect whether the quality of the remainder of message 3/13 is impaired by the double talk. 'Double talk' also occurs at step 211 when device 22 (message 15) interrupts message 6; this time it is the start of the new message 6 which is impaired. Finally, the network delay and resulting double-talk behaviour of the network causes Smith's 'goodbye Jones' (message 8: step 114) to be lost; Smith hangs up (step 117) on hearing the expected 'goodbye Smith' (step 215), but Jones only hears the line go dead (step 210). As recovery is no longer an option, Jones also hangs up (step 217).

As well as measurements of speech quality and loss during the conversation, both devices are able to compute other characteristics of the channel. Talker echo, and the operation of echo cancellers on the line, is calculated by comparing the spoken and received signals when each device is talking. Delay can be calculated from the increase in the time taken for a reply to be heard over the time for an instantaneous connection; additionally, with synchronised clocks, Jones is able to measure the one-way delay from speaking to reception using the time as spoken by Smith, though the processing of this would take place after the conversation has ended.

**CLAIMS**

1. A method of testing a communications system in which a first measurement device makes a call, through the system it is testing, to at least one  
5 second measurement device, and the devices converse using predetermined speech signals, and at least one of the devices performs measurements on the sounds it receives from the other device or devices, to compute one or more parameters describing aspects of the received signal quality, wherein the predetermined speech signals generated by each device are generated dynamically  
10 in response to the signals received by said devices, to simulate aspects of conversation over a non-perfect communications system link.
2. A method according to claim 1, in which one of the devices responds to failure to clearly receive an expected signal from another of the devices by  
15 transmitting a predetermined reply signal, intended to cause the other device to repeat the expected signal.
3. A method according to claims 1 or 2, wherein at least some of the signals to be generated for transmission in response to the input signals are selected  
20 according to the quality of the received signals.
4. A method according to claim 3, wherein at least one of the response signals selectable is a signal requesting repetition of a received signal, selectable in response to poor quality of the received signal.  
25
5. A method according to claim 3 or 4, wherein the properties of the response signals to be generated are selected according to the quality of the received signals.
- 30 6. A method according to claim 3, 4, or 5 wherein the quality of the communications system link is determined in accordance with the progress of the simulated conversation.

7. A method according to claim 6, wherein the quality of the communications system link is determined in accordance with the information content and/or quality of the response signals received.
- 5 8. A method according to any of claims 1 to 7, wherein the devices transmit and recognise device-specific identification signals forming part of the speech generated.
9. A method according to any of claims 1 to 8, including a system delay  
10 calculation process in which a first signal is transmitted from a first device to a second device, the second device receives the first signal and, after a predetermined delay, transmits a response signal to the first device, and the system delay is determined from the time of transmission of the first signal, the time of receipt of the second signal, and the predetermined delay.
- 15 10. A method according to any of claims 1 to 9, including a system delay calculation process in which two or more of the devices have synchronised clocks, wherein one of the devices transmits to another device a signal indicative of the current time, and the delay in the system under test is determined by comparing  
20 the received time-indicative signal with the time of receipt.
11. A method according to any of claims 1 to 10, wherein the signals used are pre-recorded samples of human speech.
- 25 12. A method according to any of claims 1 to 10, wherein the signals used are artificial samples of human speech.
13. A method according to claim 11 or 12, wherein the signals used are representative of a plurality of different talkers.
- 30 14. A method according to any preceding claim, in which the measurements made are used to control the operation of the communications system.

15. A method according to any preceding claim, in which one or more of the devices transmit a signal simultaneously with another signal being received by the same device, such that corruption of one or other of the signals at its respective point of receipt can be detected.
- 5
16. A device for simulating a conversation over a communications system, comprising:
- a store for storing a plurality of signals forming a conversation
  - a receiver for receiving signals from one or more complementary devices,
  - 10 a comparison device for comparing signals received by the receiver with signals stored in the store;
  - selection means responsive to the comparison device for selecting from the store a signal for transmission to the complementary device or devices;
  - transmission means for transmitting a signal selected by the selection
  - 15 means to the complementary device or devices.
17. A device according to claim 16, having conversation initiating means for generating an initial signal for transmission by the transmission means.
- 20 18. A device according to claim 16 or 17, having processing means for measuring the conversational quality of the communications system.
19. A device according to claim 18, wherein the processing means includes means for monitoring the operation of the selection means to assess the progress
- 25 of a conversation.
20. A device according to claim 18 or 19, wherein the processing means includes means for monitoring the quality of signals received from the complementary devices by the receiving means.
- 30
21. A device according to claim 18, 19, or 20 wherein the processing means includes means for monitoring the progress of the simulated conversation, and determining the quality of the communications system link thereby.



22. A device according to any of claims 16 to 21, wherein the selection means has means to select different signals for transmission according to the quality of the received signals.
- 5 23. A device according to claim 22, wherein the selection means has means to select a signal requesting repetition of a received signal in response to poor quality of the received signal
24. A device according to claim 20, 21, 22, or 23, wherein the selection  
10 means has means to select properties of the signal to be transmitted according to the quality of the received signal
25. A device according to any of claims 16 to 24, wherein the selection  
15 means has means to select different signals for transmission according to which of the predetermined signals is received.
26. A device according to claim 25, wherein the selection means has means  
20 for responding to a predetermined received signal by reselecting the signal previously transmitted by the transmitter.
27. A device according to any of claims 16 to 26, wherein the device has means for transmitting an identification signal as part of the speech generated.
28. A device according to any of claims 16 to 27, wherein the device has  
25 means to recognise identification signals specific to the individual member of a plurality of complementary devices transmitting the speech received by the receiver.
29. A device according to claim 27 or 28, wherein the identification signals  
30 are spoken words.

30. A device according to any of claims 16 to 29, having means for simultaneously transmitting and receiving predetermined signals, and means for detecting whether the predetermined received signal is corrupted by the presence of the transmitted signal.

5

31. A device according to any of claims 16 to 30, including delay calculation means comprising means for determining the time at which a first signal is transmitted to a complementary device, means for determining the time at which a predetermined response signal is received from the complementary device, means  
10 for retrieving from the store the value of a predetermined delay at the complementary device between receipt of the first signal and transmission of the response signal, and means for determining therefrom the additional delay imposed by the system under test.

15 32. A measurement system for assessing the quality of a communications system by simulating a conversation between human talkers, the system consisting of two or more devices connected to the system under test at appropriate points for the initiation of a call and transmission of speech; the devices having means to generate predetermined conversational patterns which  
20 vary dynamically in a way that is representative of human speech, such that the speech generated at each device varies dynamically in response to the signals received at that device, and at least one of the devices having processor means to extract a number of different quality parameters from the received signals.

25 33. A measurement system according to claim 32, comprising at least two devices each according to one of claims 16 to 31.

34. A measurement system according to claim 32 or 33, wherein two or more of the devices have synchronised clocks, wherein at least one of the devices has  
30 means for transmitting a signal indicative of the current time, and at least one other of the devices has means for determining the delay in the system under test by comparing the received time-indicative signal with the time of receipt.

35. A method of testing a communications system substantially as described, with reference to the drawings.

36. A device for simulating a conversation over a communications system,  
5 substantially as described, with reference to the drawings.

37. A measurement system for assessing the quality of a communications system substantially as described, with reference to the drawings.

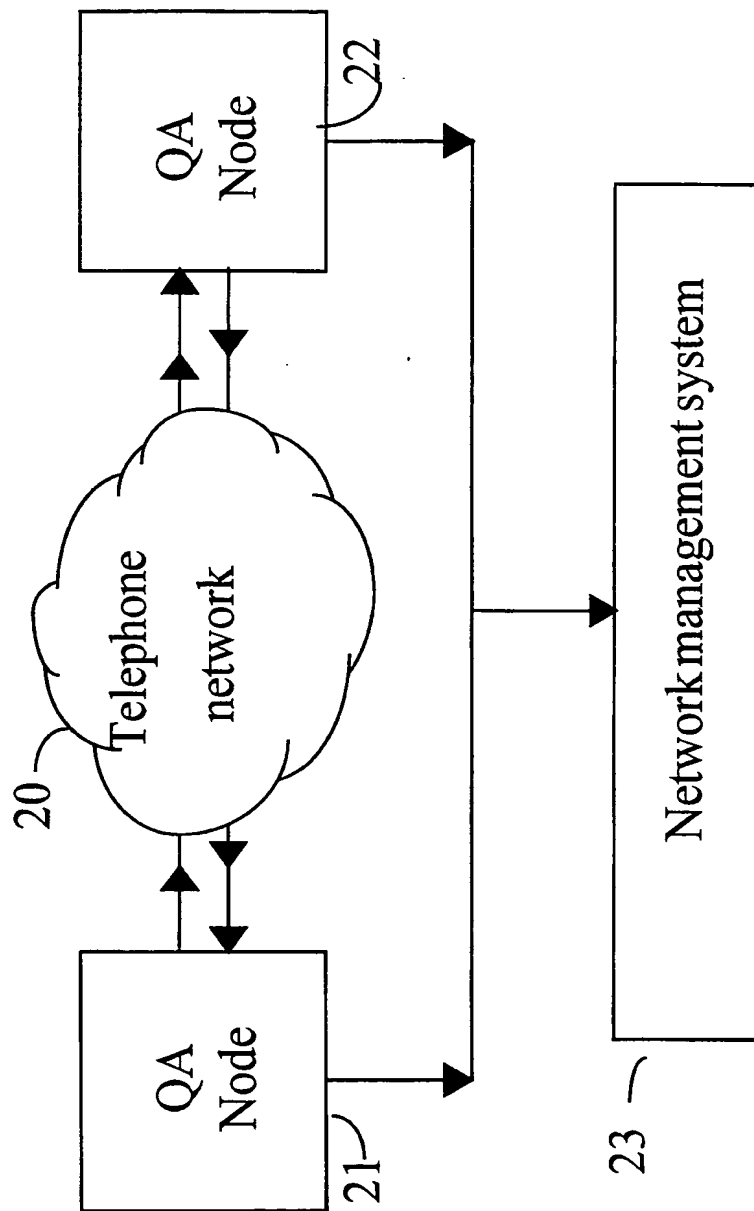


Figure 1

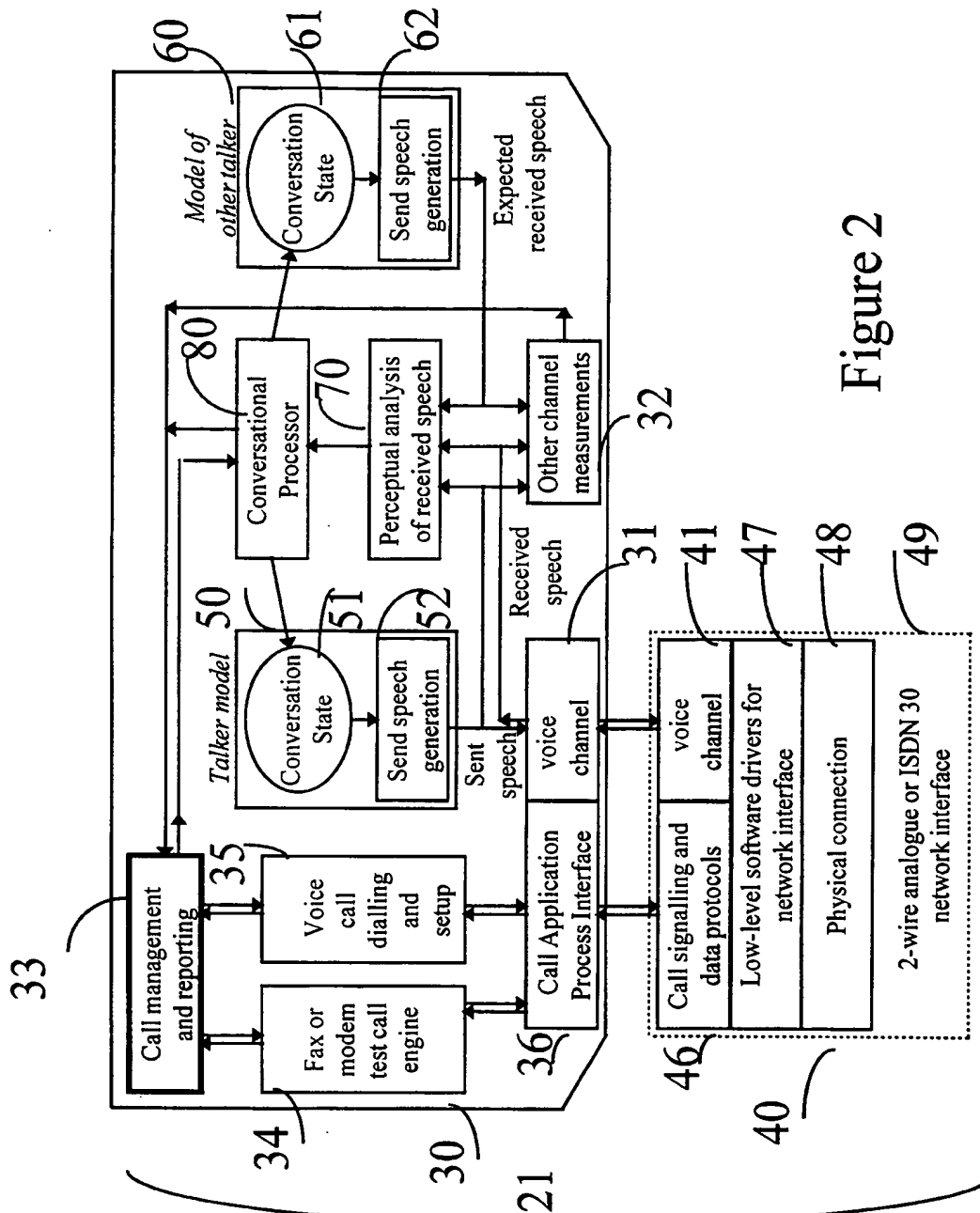
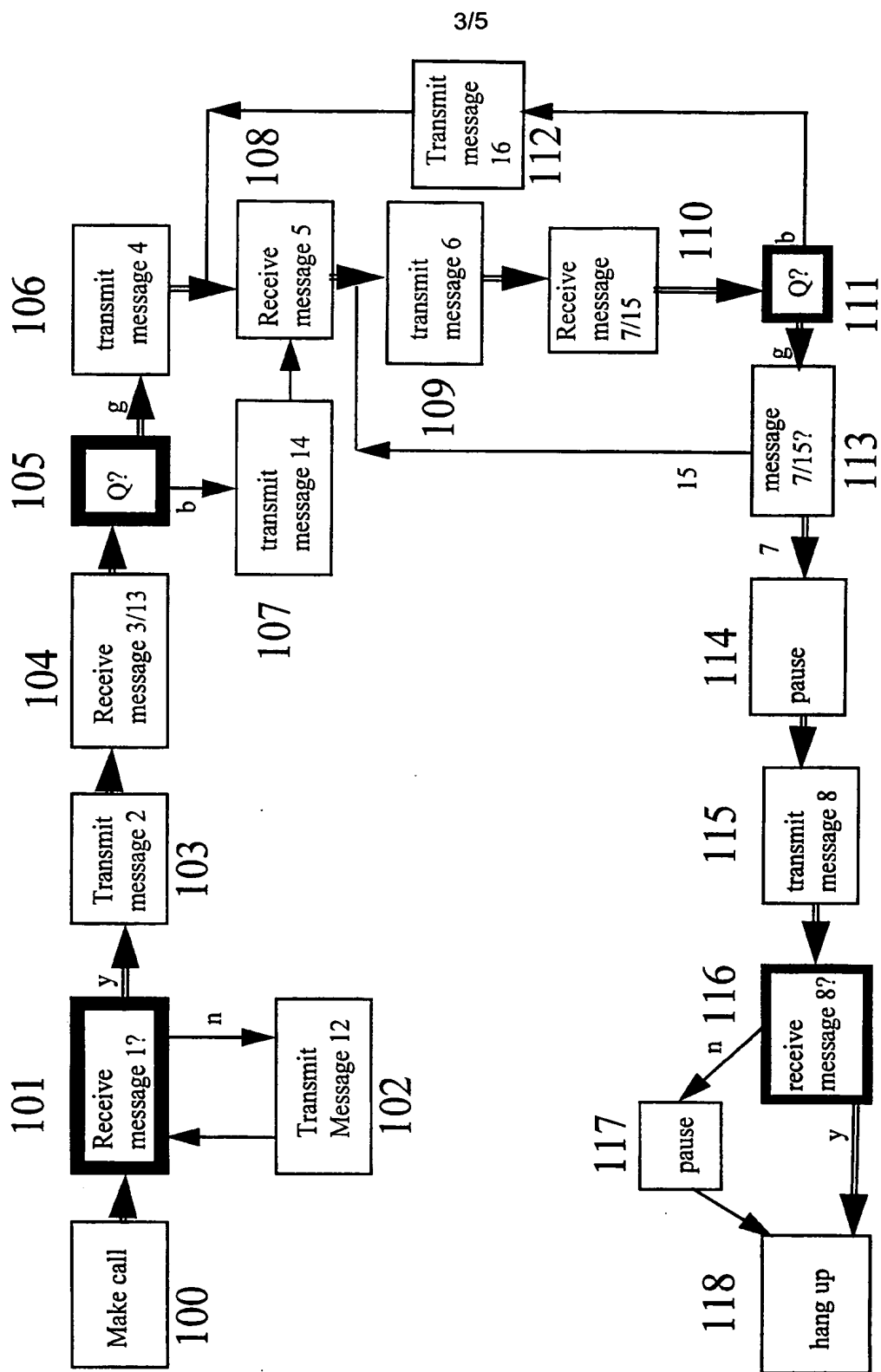


Figure 2



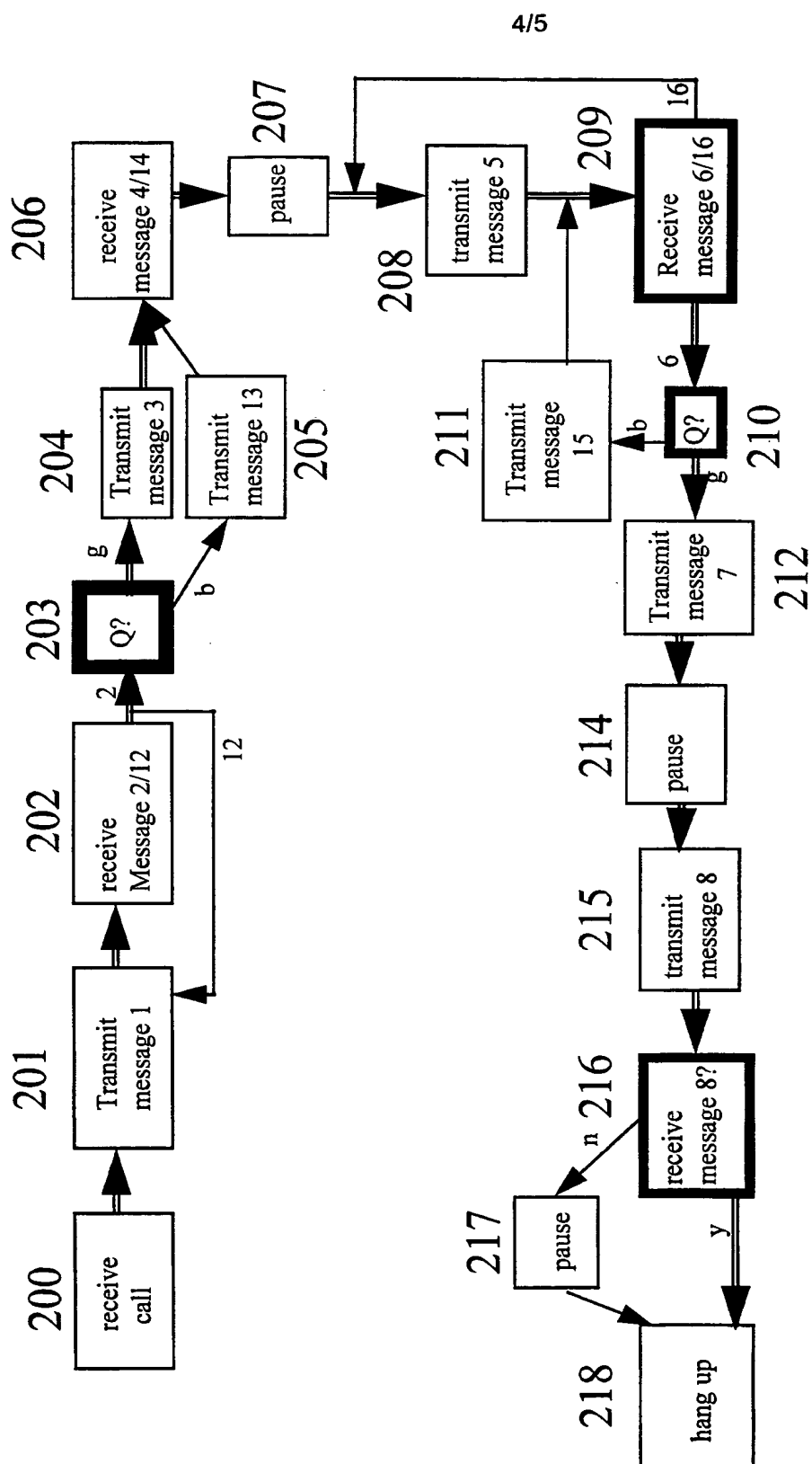


Figure 4

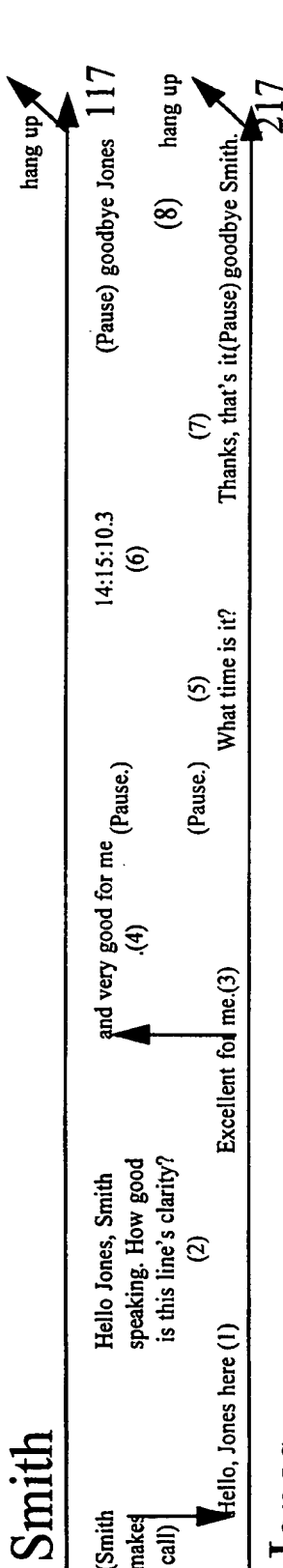


Figure 5

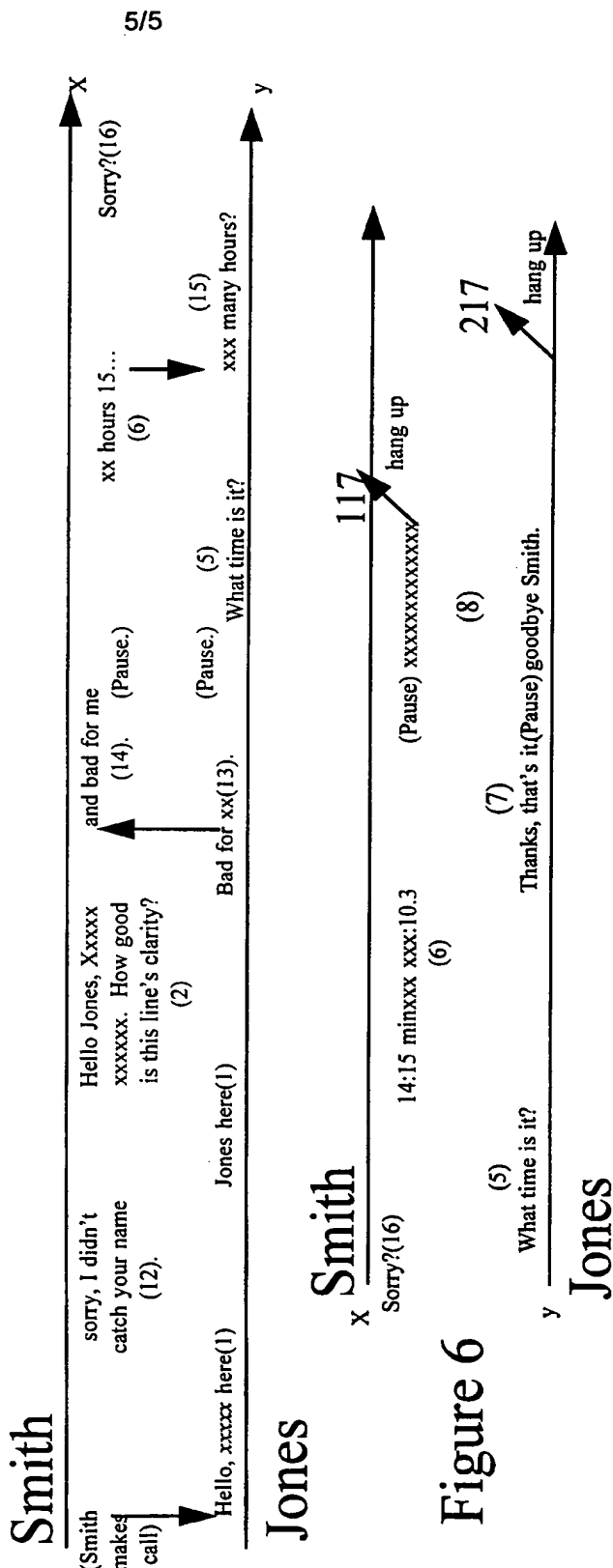


Figure 6



# INTERNATIONAL SEARCH REPORT

Inter: 1st Application No

PCT/GB 98/01305

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 H04M3/22

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 567 439 A (TELEVERKET) 27 October 1993 see page 3, line 45 - page 4, line 15 ----	1, 16, 32
A	US 5 369 644 A (OLIVER DOUGLAS W) 29 November 1994 ----	
A	EP 0 298 621 A (OLIVETTI & CO SPA) 11 January 1989 ----	
A	EP 0 722 164 A (AT & T CORP) 17 July 1996 -----	

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

16 July 1998

Date of mailing of the international search report

22/07/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Vandevenne, M

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter: nal Application No

PCT/GB 98/01305

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0567439 A	27-10-1993	SE 469007 B GB 2266432 A,B JP 6054364 A SE 9201236 A US 5742929 A	26-04-1993 27-10-1993 25-02-1994 26-04-1993 21-04-1998
US 5369644 A	29-11-1994	NONE	
EP 0298621 A	11-01-1989	DE 3886460 D DE 3886460 T JP 1049361 A JP 1958270 C JP 6093715 B US 4847885 A	03-02-1994 16-06-1994 23-02-1989 10-08-1995 16-11-1994 11-07-1989
EP 0722164 A	17-07-1996	US 5715372 A CA 2165229 A JP 8314884 A SG 44395 A US 5758027 A	03-02-1998 11-07-1996 29-11-1996 19-12-1997 26-05-1998